PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : A61K 31/70, 31/715

A2

(11) International Publication Number: WO 93/01820

(43) International Publication Date: 4 February 1993 (04.02.93)

(21) International Application Number: PCT/US92/05985

(22) International Filing Date: 16 July 1992 (16.07.92)

(30) Priority data:
731,214 16 July 1991 (16.07.91) US

(71) Applicant: BRISTOL-MYERS SQUIBB COMPANY [US/US]; 345 Park Avenue, New York, NY 10154 (US).

(72) Inventor: CURTIS, Benson, M. ; 1520 N.W. Woodbine Way, Seattle, WA 98177 (US).

(74) Agents: POOR, Brian, W.; Bristol-Myers Squibb Company, 3005 First Avenue, Seattle, WA 98121 (US) et al.

(81) Designated States: AU, CA, FI, JP, KR, NO, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LU, MC, NL, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: INHIBITION OF NON-CD4 MEDIATED HIV INFECTION

(57) Abstract

A specific nonCD4-gp120 receptor has been isolated which has specific binding affinity for the gp120 surface protein of human immunodeficiency virus (HIV). Methods of treating HIV infection of CD4 negative cells, such as colon and brain, are disclosed together with methods of detecting HIV and diagnostic kits.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	ML	Mali
AU	Australia	FR	France	MN	Mongolia
BB	Barbados	GA	Gabon	MR	Mauritania
BE	Belgium	GB	United Kingdom	MW	Malawi
BF	Burkina Faso	GN	Guinea	NL	Netherlands
BG	Bulgaria	GR	Greece	NO	Norway
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IE	Ireland	RO	Romania
CA	Canada	IT	Italy	RU	Russian Federation
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		
ES	Spain	MG	Madagascar		

INHIBITION OF NON-CD4 MEDIATED HIV INFECTION

TECHNICAL FIELD OF THE INVENTION

The present invention is directed to a non-CD4 cell surface receptor for gp120. This gp120 receptor (gp120r) has been isolated and cloned and is utilized in the present invention in methods and kits for the inhibition and detection of HIV infection.

BACKGROUND OF THE INVENTION

Two types of human retroviruses have been identified, leukemia viruses and AIDS-related viruses. The primary targets of the human retroviruses are T lymphocytes and cells of the central nervous system. All human retroviruses are transmitted by intimate contact, blood contamination, and infection in utero or after birth by milk. It is likely that all human retroviruses originated in Africa and that they encountered the human species via interspecies infection, possibly from African green monkeys or a related species. The human retroviruses first discovered, Human T Lymphotropic Virus Type 1 (HTLV-1) and Human T Lymphotropic Virus Type II (HTLV-II), have a preferential tropism for T4 cells and some T8 cells, share significant sequence homology, and are mainly associated with T cell leukemias and lymphomas. The other group of human retroviruses, generally called Human Immunodeficiency Viruses (HIV), is discussed in greater detail below. There are two major differences between the two types of human retroviruses: (1) there is substantial genomic variability among various HIV isolates, whereas the genomes of HTLV-I and HTLV-II are stable; and (2) HIV entered human populations much more recently than HTLV-I or HTLV-II.

The human immunodeficiency virus (HIV) is a cytopathic retrovirus and the causative agent of the acquired immunodeficiency syndrome (AIDS). Two forms of HIV have now been identified. The prototype virus, HIV-1, previously termed lymphadenopathy-associated virus (LAV) and Human T Lymphotropic Virus Type III (HTLV-III), is responsible for the vast majority of reported AIDS cases worldwide. Another retrovirus, HIV-2, has been isolated primarily from West African patients with AIDS and is pathogenically related to HIV-1. On the genetic level, HIV-2 is actually more closely related to the simian immunodeficiency virus (SIV), a retrovirus infecting monkeys.

Over half of the people that have contracted AIDS in the United States have already died. As many as three million persons in this country may be asymptomatic carriers of

HIV and are capable of transmitting the virus. It had been estimated in 1986 that 270,000 cases of AIDS will have occurred in the United States by 1991 (U.S. Public Health Service, (1986), Public Health Rep. 101:341). The mortality rate from AIDS is disturbingly high, exceeding 80% within three years of diagnosis and possibly reaching 100% over a longer period.

Worldwide, the AIDS epidemic may involve some five to ten million presently infected persons. Particularly troublesome are statistics from the African continent where millions of individuals are believed infected with HIV, deaths range in the hundreds of thousands, and heterosexual transmission predominates. To date, there is neither a known cure for AIDS nor an effective vaccine against HIV infection.

HIV is a member of the nontransforming, cytopathic lentivirus family of retroviruses. HIV causes a typically fatal disease characterized by severe immunodeficiency or neurodegenerative disease, or both. The primary basis for HIV induced immunosuppression is the depletion of the helper/inducer subset of T lymphocytes expressing the CD4 molecule (T4 or $CD4^+$ cells), which serves as a high affinity cell surface receptor for the virus. T4 lymphocytes are involved directly or indirectly in the induction of nearly every immunologic function in the body, and their depletion results in susceptibility to a wide range of opportunistic infections and neoplasms.

In addition to the T4 lymphocyte, other cells expressing the CD4 molecule are targets of HIV infection, especially monocyte-macrophages. HIV infection also results in serious B cell abnormalities including polyclonal activation, hypergammaglobulinemia, elevated levels of circulating immune complexes, and autoantibodies. A decreased number of functional natural killer (NK) cells have also been observed in AIDS patients.

Infection of $CD4^+$ cells is initiated by the interaction of the CD4 molecule with the major HIV envelope glycoprotein gp120, an event which is followed by internalization and uncoating of the virion, transcription of genomic RNA to DNA by virus-encoded reverse transcriptase, and integration of the resulting proviral DNA into host cell chromosomal DNA. Also, unintegrated proviral DNA accumulates in large amounts within infected cells and is probably a significant factor in HIV cytopathology (Shaw et al., (1984) Science 226:1165).

The depletion of $CD4^+$ T cells appears to contribute significantly to the immunosuppression associated with AIDS. A primary cytopathic effect of the virus in vitro is HIV-induced syncytium formation. CD4, through its interaction with gp120 plays an important role in syncytium formation. However, it has been observed that molecules on the cell surface of uninfected cells other than CD4 are also involved in HIV-induced cell fusion (Hildreth et al. (1989) Science 244:1075-1078).

Infection by HIV produces, in addition to AIDS, a set of neuropsychiatric disorders which are called the AIDS dementia complex (ADC) (Price et al., (1988) 239:586-592).

The symptoms of ADC include cognitive impairment, apathy and motor dysfunctions, and may affect as many as 90% of AIDS victims. The underlying cause of ADC appears to be the death of brain cells and HIV-1 can be isolated from the brains of infected individuals (Ho et al, (1987) N. Eng. J. Med. 317:278-286).

An early study suggested that the cellular attachment site for HIV in brain might be CD4 (Pert et al., (1986) Proc. Natl. Acad. Sci. USA 83:9254-9258) but attempts to replicate these findings were not successful (Kozlowski et al., (1989) Neurosci. Abstr. 15:671). It now appears unlikely that the CD4 antigen is involved in the infection of brain-derived cells by HIV. Susceptibility of brain cells to infection with HIV-1 does not correlate with the level of expression of CD4 (Chang-Mayer et al., (1987) Proc. Natl. Acad. Sci. USA 84:3526-3530; Srinivasan et al., (1988) Arch. Virol. 98:135-141), and infection of brain-derived cells by HIV-1 is not blocked by anti-CD4 antibodies (Clapham et al., (1989) Nature 337:368-370; Li et al., (1990) J. Virol. 64:1383-1387).

The present invention demonstrates the presence of a non-CD4 receptor for gp120 and a method for the inhibition of HIV infection of cells such as brain and muscle which do not express high levels of CD4.

SUMMARY OF THE INVENTION

Many cells that are susceptible to HIV infection appear to bind gp120 through a non-CD4 surface protein. The present invention has identified this non-CD4 gp120 receptor (gp120r) and has recombinantly expressed and characterized gp120r.

In this invention a specific non-CD4 gp120r has been isolated which has specific binding activity for gp120 present on Human Immunodeficiency Virus-1 (HIV). This gp120r has a molecular weight of about 45, 000 daltons, contains about 400 amino acid residues and is characterized by a Kd for gp120 of about 1.3 nM to about 2.0 nM. The binding of gp120 to gp120r is inhibited by specific carbohydrates, such as mannose and fucose, plant lectins such as concanavalin A and specific antibiotics, such as pradimicin A.

In one embodiment of the present invention, a cDNA molecule that transcribes an mRNA encoding for gp120r is cloned and expressed to produce gp120r. The DNA is selected from a gene library obtained from tissue such as placenta, brain, muscle and colon.

A method of inhibiting HIV infection of mammalian cells, such as brain, muscle and neural cells, is contemplated by the present invention. In this method, cells are contacted with an effective amount of an appropriate inhibitor of gp120r binding for a time

period sufficient to significantly inhibit the binding of HIV to the non-CD4 protein, gp120r. Specific inhibitors of gp120r binding include mannose carbohydrates, fucose carbohydrates, plant lectins, and antibiotics such as pradimicin A.

The gp120r of the present invention can also be utilized in a method and a kit for the detection of the presence of HIV in a fluid sample. In this method, the binding of HIV to gp120r is detected by an indicating means such as a labelled antibody capable of binding to the HIV-gp120r reaction product. It is also contemplated that the gp120r can be affixed to a solid matrix to form a solid support that is useful in this method and/or kit.

DESCRIPTION OF THE FIGURES

In the drawings:

FIGURE 1 illustrates expression cloning of the gp120r cDNA and comparison to CD4.

A: Autoradiography of gp120 binding to gp120r and CD4 expressed in COS cells. A-F [^{125}I]vgp120; A, gp120r; B, gp120r with G17-2; C, gp120r with 200 nM unlabelled bgp120; D, CD4; E, CD4 with G17-2; F, CD4 with bgp120. G-L [^{125}I]ngp120; G, gp102r; H, gp120r with 110.1; I, gp120r with bgp120; J, CD4: k, CD4 with 110.1; L, CD4 with bgp120.

B: Inhibition of [^{125}I]vgp120 binding to gp120r and CD4. A-F gp120r and G-L CD4. A+G, HIV antisera (1:20; Trimar); B+H, D-galactose (100 mM); C+I, D-mannose (100 mM); D+J, L-fucose (100 mM); E+K, Concanavalin A (1 mg/ml); F+L, pradimicin A (100 μg/ml).

C: gp120r binding of HIV. A, HIV; B, HIV with 200 nM bgp120.

FIGURE 2 illustrates the characterization of the gp120r.

A: Scatchard analysis of [^{125}I]gp120 binding. ▲ - ▲, vgp120 binding to placenta, Kd 1.3 nM, B_{max} 19 fmol/mg protein; ■ with μg/ml G17-2; ● - ●, vgp120 binding to gp120r COS cells, Kd 1.7 nM, B_{max} 150,000 receptors/cell (R/C); ○, ngp120, Kd 1.8 nM, 149,000 R/C.

B: Inhibition of [^{125}I]gp120 binding to gp120r COS cells. Open symbols ngp120, filled symbols vgp120. The relative values were the same with both forms of gp120. Mannan expressed as mg/ml. ❑, mannan (IC50 6 μg/ml); ●, L-fucose (K_i 6 mM); Δ, α-methyl D-mannoside (K_i 15 mM), ○, D-mannose (K_i 23 mM); ◇, N-acetylglucosamine (K_i 70 mM), ■, EGTA (K_i 0.3 mM).

C: Internalization of gp120 by gp120r COS cells. Points represent the mean of two experiments with vgp120 and ngp120. ● - ●, suface; ○ - ○ internal.

D: Placenta control sera; 2, placenta HIV sera; 3, gp120r COS control sera; 4, gp120r COS HIV sera.

E: Northern blot of gp120r expression. Polyadenylated (A^+); 2, placenta; 3, thymus; 4+12, forebrain; 5, skeletal muscle; 6, heart; 7, liver; 8, kidney; 9, colon; 10 medulla; 11, cerebellum; 13, T cell (CEM; 16 μg A^+) 14, B cell (TS-1; 16 μg A^+); 15, macrophage (U937; 8 μg A^+); 16, cervical carcinoma (HeLa; 16 μg A^+). The different apparent size of the ~5 kb band is an artifact of displacement by 28S rRNA.

FIGURE 3 illustrates the sequence analysis of the gp120r.

A: Nucleotide and deduced protein sequence of gp120r cDNA.

B: Hydropathicity plot of the gp120r. The predicted transmembrane segment and the start of the eight amphipathic repeats are indicated by arrows.

C: Aminoacid alignment of the gp120r C-type lectin domain.

DESCRIPTION OF PREFERRED EMBODIMENTS

HIV infection of brain and muscle cell lines is not blocked by soluble CD4 or anti-CD4 antibodies (Clapham, P.R. et al., (1989) Nature 337:368-370; Harouse, J.M. et al., (1989) J. Virol. 63:2527-2533; Weber, J. et al., (1989) J. Gen. Virol. 70:2653-2660). This is consistent with the existence of a second gp120 receptor. Binding studies indicated that human placenta was another source for a non-CD4 gp120 receptor, and a cDNA for a second gp120 receptor (gp120r) was isolated by the present invention from a placental library. The gp120r has a higher binding affinity for gp120 than CD4. Sequence analysis revealed homology to membrane associated C-type lectins, and inhibition studies have shown that the receptor binds gp120 through a mannose or fucose containing carbohydrate. The gp120r rapidly internalizes gp120, and is expressed in placenta, thymus, muscle, and colon. These results, when considered with previous studies on the role of gp120 carbohydrate in HIV infection (Lifson, J. et al., (1986) J. Exp. Med. 164:2101-2106; Ezekowitz, R.A.B. et al., (1989) J. Exp. Med. 169:185-196; Larkin M. et al., (1989) AIDS 3: 793-798; Tanabe-Tochikura A. et al., (1990) Virology 176:473-476), suggest a potential role for the gp120r in HIV infection or pathology.

The present invention demonstrates that the gp120r participates in cellular binding of HIV by a non-CD4 pathway in muscle and brain, as well as, facilitating virus attachment in CD4 positive cell types. It is likely that the gp120r plays a significant role in

transplacental transport of HIV (Zacher, V. et al., (1991) J. Virol. 65:2102-2107) and colon infection (Barnett, S.W. et al. (1991) Virol. 182:802-809). Gp120 produces an increase in intracellular calcium in rat retinal ganglion cells (Dreyer, E.B. et al., (1990) Science 248:364-367) suggesting that the gp120r or a homologous protein may have signaling functions in the nervous system disrupted by gp120 leading to HIV neurotoxicity.

In the present invention, a new non-CD4 binding protein, or receptor, for gp120 was isolated. The HIV surface protein gp120 was found to bind to a receptor on human placental membranes that was not blocked by antibodies directed against CD4, such as G17-2 and OKT4a, and which interfere with gp120 binding to CD4. A cDNA encoding this receptor was isolated from a placental cDNA library in a mammalian expression vector (pCDM8). The gene products were expressed in COS cells and were screened by ^{125}I-labelled gp120 binding. From a pool of 90,000 cDNA molecules, a single clone was isolated that encoded a protein which bound gp120, even in the presence of concentrations of anti-CD4 antibody (G17-2) which completely blocked gp120 binding to CD4.

Sequence studies were carried out and indicated that the 1.5 kilobase cDNA clone encoded a previously unknown member of a family of Type II membrane proteins with an extracellular C type lectin domain.

The cloned gp120r of the present invention binds gp120 with an affinity (Kd) of about 1 to 2 nM, which is considerably greater than the affinity of CD4 for gp120 (about Kd = 4 nM).

The binding of gp120 to gp120r is not blocked by polyclonal HIV antisera, but is inhibited by mannose carbohydrates, fucose carbohydrates, plant lectins such as concanavalin A and pradimicin A antibiotics. Other sugars such as N-acetyl-d-glucosamine and galactose are less potent inhibitors.

The gp120r is expressed on many mammalian cells which do not exhibit high levels of CD4, such as placenta, skeletal muscle, brain, and mucosal cells. Other tissue and cells displaying gp120r include colon, thymus, heart, T cells, B cells and macrophages. The distribution of tissue having gp120r parallels that for binding of gp120 which is not blocked by CD4 antibodies, and for HIV infection which is not neutralized by soluble CD4. This observation suggests a role for gp120r in viral infection.

In gp120r expressing transfected COS cells, gp120 is rapidly internalized following binding to gp120r. This binding and internalization of gp120 is inhibited by compounds such as mannan, concanavalin A and pradimicin A.

In the present invention a cDNA which encodes gp120 was isolated and cloned. A DNA molecule of the present invention corresponds to a complementary DNA molecule which transcribes a messenger RNA (mRNA) molecule which, when translated, encodes gp120r. The cDNA molecules were obtained by reverse-transcribing mRNA molecules

isolated from mammalian tissue such as placenta, colon, brain or thymus. The transcription and cloning of cDNA molecules and isolation of gene products are techniques well known in the art and, for example, are described in Sambrook et al., "Molecular Cloning: A Laboratory Manual", 2d edition, Cold Spring Harbor Lab., Cold Spring Harbor, NY (1989), which is incorporated herein by reference.

As used herein, the phrases "physiologically tolerable" and "pharmaceutically acceptable" refer to molecular entities and compositions that do not produce an allergic or similar untoward reaction, such as gastric upset, dizziness and the like, when administered to a mammal. The physiologically tolerable carrier may take a wide variety of forms depending upon the preparation desired for administration and the intended route of administration.

A carrier is a material useful for administering the active compound and must be "acceptable" in the sense of being compatible with the other ingredients of the composition and not deleterious to the recipient thereof.

The pharmaceutical compositions are prepared by any of the methods well known in the art of pharmacy all of which involve bringing into association the active compound and the carrier therefor.

For therapeutic use, the agent utilized in the present invention can be administered in the form of conventional pharmaceutical compositions. Such compositions can be formulated so as to be suitable for oral or parenteral administration, or as suppositories. In these compositions, the agent is typically dissolved or dispersed in a physiologically tolerable carrier.

As an example, the compounds of the present invention can be utilized in liquid compositions such as sterile suspensions or solutions, or as isotonic preparations containing suitable preservatives. Particularly well suited for the present purposes are injectable media constituted by aqueous injectable isotonic and sterile saline or glucose solutions. Additional liquid forms in which the present compounds may be incorporated for administration include flavored emulsions with edible oils such as cottonseed oil, sesame oil, coconut oil, peanut oil, and the like, as well as elixirs and similar pharmaceutical vehicles.

The present agents can also be administered in the form of liposomes. As is known in the art, liposomes are generally derived from phospholipids or other lipid substances. Liposomes are formed by mono- or multi-lamellar hydrated liquid crystals that are dispersed in an aqueous medium. Any non-toxic, physiologically acceptable and metabolizable lipid capable of forming liposomes can be used. The present compositions in liposome form can contain, in addition to the agent of the present invention, stabilizers,

preservatives, expedients, and the like. The preferred lipids are the phospholipids and the phosphatidyl cholines (lecithins), both natural and synthetic.

Methods to form liposomes are known in the art. See, for example, Prescott, Ed., "Methods in Cell Biology", Volume XIV, Academic Press, New York, N.Y. (1976) p 33 et seq.

The present compounds can also be used in compositions such as tablets or pills, preferably containing a unit dose of the compound. To this end, the agent (active ingredient) is mixed with conventional tabletting ingredients such as corn starch, lactose, sucrose, sorbitol, talc, stearic acid, magnesium stearate, dicalcium phosphate, gums or similar materials as non-toxic, physiologically tolerable carriers. The tablets or pills of the present compositions can be laminated or otherwise compounded to provide unit dosage forms affording prolonged or delayed action.

It should be understood that in addition to the aforementioned carrier ingredients the pharmaceutical formulation described herein can include, as appropriate, one or more additional carrier ingredients such as diluents, buffers, flavoring agents, binders, surface active agents, thickeners, lubricants, preservatives (including antioxidants) and the like, and substances included for the purpose of rendering the formulation isotonic with the blood of the intended recipient.

The tablets or pills can also be provided with an enteric layer in the form of an envelope that serves to resist disintegration in the stomach and permits the active ingredient to pass intact into the duodenum or to be delayed in release. A variety of materials can be used for such enteric layers or coatings, including polymeric acids or mixtures of such acids with such materials as shellac, shellac and cetyl alcohol, cellulose acetate, and the like. A particularly suitable enteric coating comprises a styrene-maleic acid copolymer together with known materials that contribute to the enteric properties of the coating.

A method of inhibiting HIV infection of mammalian cells is disclosed in the present invention. A pharmaceutical composition containing a compound which effectively inhibits the binding of gp120r to HIV, is contacted with cells either in vitro or in vivo for a time period sufficient to significantly inhibit the binding of HIV to the cell surface.

Compounds effective in this method include mannose carbohydrates, fucose carbohydrates, plant lectins and pradimicin A antibiotics. Specifically preferred compounds are mannose, fucose, mannan, concanavalin A and pradimicin A. The pharmaceutical composition of the present invention includes a compound which effectively inhibits gp120r binding to HIV and may also include a physiologically tolerable carrier.

The method of the present invention is preferably utilized to inhibit HIV infection of placental, brain, muscle, neural and colon cells.

A diagnostic method is also described in the present invention for detecting the presence, and preferably the amount, of HIV present in a fluid sample by producing a reaction product containing HIV bound to gp120r. Those skilled in the art will recognize that there are well known clinical diagnostic procedures that can be utilized for the formulation and detection of such reaction products. Thus, while exemplary assay methods are described herein, the invention is not intended to be so limited.

Various heterogeneous and homogeneous assay protocols can be employed for detecting the presence, and preferably the amount, of HIV in a fluid sample. For example, the present invention contemplates a method for assaying a sample, such as a body fluid, for the presence of HIV comprising the steps of:

(a) admixing a fluid sample with gp120r, either in solution or affixed to a solid matrix;

(b) maintaining the admixture for a predetermined time period such as about 10 minutes to about 16 - 20 hours and under biological assay conditions at a temperature of about 4°C to about 45°C that is sufficient for any HIV present in the sample to react with (bind) the gp120r to form a reaction product; and

(c) determining the presence of any reaction product that is formed, and thereby the presence of any HIV in the admixture.

Preferably, the fluid sample is a body fluid sample, such as blood, plasma, serum, urine, saliva, semen or cerebrospinal fluid (CSF).

The determination of the presence of a reaction product, either directly or indirectly, can be accomplished by assay techniques well known in the art such as by the use of an indicating or labelling means, as discussed hereinbelow. In a preferred embodiment, a labelled indicating means, such as a fluorescein-labelled antibody, is capable of binding to the gp120r present in the reaction product to form a labelled complex. Determining the presence of the labelled complex provides an assay for the presence of HIV in the sample. In particularly preferred embodiments, the amount of labelled indicating means bound as part of the complex is determined, and thereby the amount of HIV present in the sample is determined. When that amount is zero, no HIV is present in the sample, within the limits of detection. Methods for assaying the presence and amount of a labelled indicating means depend on the label used, such labels and assay methods being well known in the art.

In a preferred embodiment, the gp120r is affixed on a solid matrix to form a solid phase support. In that embodiment, the assay is heterogeneous, solid/liquid phase assay and, as such, has its own preferred manipulations. For example, following admixing of a liquid sample with a solid support containing gp120r affixed thereto, the admixture is

also a detection means. In one embodiment, the package can contain a microtiter plate well to which microgram quantities of gp120r have been operatively affixed, ie., linked so as to be capable of reacting with and bind HIV and/or gp120.

As used herein, the terms "label" "indicating means" and "labelled indicating means", in their various grammatical forms refer to single atoms and molecules that are either directly or indirectly involved in the production of a detectable signal to indicate or detect the presence of a reaction product. Such labels are themselves well known in clinical diagnostic chemistry and constitute a part of this invention only insofar as they are utilized with otherwise novel methods and/or systems.

The indicating means can be a fluorescent labelling agent that chemically binds to antibodies or protein antigens without denaturing them to form a fluorochrome (dye) that is a useful immunofluorescent tracer. Suitable fluorescent labelling agents are fluorochrome, such as fluorescein isocyanate (FIC), fluorescein isothiocyanate (FITC), 5-dimethylamine-1-naphthalene sulfonyl chloride (DANSC), tetramethylrhodamine isocyanate (TRITC), lissamine and the like. Immunofluorescence analysis techniques are well known in the art, and for example, is described in DeLuca, "Immunofluorescence Analysis" in <u>Immunofluorescence Analysis</u>, Marchalonis et al., (1982) eds., John Wiley & Sons, Ltd., pp. 189-231, which is incorporated herein by reference.

Other preferred indicating means are colorimetric agents and enzymes, such as horseradish peroxidase, glucose oxidase or the like, linked as described above, as well as radioactive elements, preferably an element that produces gamma ray emissions. Elements which emit gamma rays, such as ^{124}I, ^{125}I, ^{128}I, ^{132}I, and ^{51}Cr represent one class of radioactive indicating groups. Another group of useful labelling means are those elements such as ^{11}C, ^{18}F, ^{15}O and ^{13}N which emit positrons. The positrons so emitted produce gamma rays upon interaction with electrons present.

Having generally described this invention, a further understanding can be obtained by reference to certain specific examples which are provided herein for purposes of illustration only and are not intended to be limiting unless otherwise specified.

EXAMPLE 1

<u>Cloning and Isolation of Non-CD4 Gp140 Receptor Protein</u>

Human placental membranes were found to be able to bind vaccinia derived recombinant gp120 (vgp120) with a Kd of 1.3 nM. At nM (concentrations) of gp120 none of this binding was inhibited by an antibody (G17-2) which has been reported to efficiently block gp120 binding to CD4 (Linsley et al. (1988) J. Virol. <u>62</u>:3695-3702), as shown in

FIGURE 2A. Approximately 50 - 90% of the total placental gp120 binding was not due to CD4.

A placental cDNA library was obtained in the mammalian expression vector pCDM8 and was screened. A cDNA was isolated which expressed protein that exhibited high affinity binding for vgp120 in the presence of G17-2.

This protein, designated as gp120 receptor (gp120r), also bound native gp120 (ngp120), and the binding component was precipitated in the presence of an antibody directed against gp120.

EXAMPLE 2

Characterization

The binding of radiolabelled gp120 to gp120r expressed in COS-7 cells was studied. Pools of 90,000 cDNA molecules, obtained from a placental pCDM8 library, were transfected by electroporation into COS-7 cells. Cells which expressed gp120r on the surface was identified by screening with either 1 nM of ^{125}I-labelled vgp120 (^{125}I-vgp120) or ^{125}I-ngp120 by the method described in Kozlowski et al., (1990) Antivir. Chem. Chemother. 1:175-182, incorporated herein by reference. The results of binding studies utilizing the transfected COS-7 cells are shown in FIGURE 1.

Binding of labelled gp120 (1 nM) to the cells was carried out following a 1 hour preincubation of the cells or GP120 at 22°C with one or more of the following: anti-CD4 antibody G17-2 (5 ug/ml), baculovirus-derived gp120 (bgp120, American Biotechnologies, 200 nM), anti-gp120 monoclonal antibody 110.1 (25 μg/ml), D-mannose (100 mM), D-galactose (100 mM), L-fucose (100 mM), concanavalin A (1 mg/ml) or pradimicin A (100 ug/ml). The cells were monitored after autoradiography (3 days). The results seen in FIGURES 1 (A and B) illustrate that gp120 binding to the gp120r expressed on the cells was blocked by excess bgp120, mannose, fucose, pradimicin A, Concanavalin A, and preincubation with antibody 110.1 but not by CD4, antibody G17-2, galactose, or HIV antisera. Studies were also carried out on gp120 binding to CD4 expressing COS cells, transfected with π H3MCD4 by the method of Peterson et al. (1988) Cell 54:65-72.

Control studies of the binding of ^{125}I-labelled psoralen-UV inactivated HIV-BRU to the gp120r expressing COS-7 cells demonstrated binding of HIV to gp120r and blockage by excess bgp120 (FIGURE 1C). A tabular compilation quantitating the amount of bound material to the cells in FIGURE 1 is shown in Table. 1

TABLE 1

FIGURE 1	LABELLED MATERIAL	COMPETITION WITH	CPM BOUND X 10^{-3} GP120R	CD4
A	vgp120+	—	60	20
		G17-2	60	5
		bgp120	6	4
B	vgp120+	HIV antisera	63	3
		D-galactose	50	20
		D-mannose	6	20
		L-fucose	6	20
		concanavalin A	8	6
		pradimicin A	8	6
		OKT4A	60	5
		N-acetylgalactosamine	60	20
		N-acetylglucosamine	30	20
		mannan	6	20
		mannose-6-phosphate	60	20
		sialic acid	60	20
		human IgE	60	20
C	HIV-BRU+	—	8	4
		bgp120	2	2

Scatchard plots of gp120 binding to placental membranes and to COS cells expressing the gp120r were carried out in the presence and absence of a 200 fold excess of bgp120 or ngp120. The results, shown in FIGURE 2A, demonstrate a specific binding of vgp120 to gp120r with a Kd of 1.7 nM ± 0.4 (n=4) and of ngp120 to gp120r with Kd of 1.8 nM ± 0.2 (n=4), with 150,000 and 149,000 receptors per cell, respectively. Concurrent analysis of gp120 binding to CD4 expressed on COS cells gave a Kd of 4-5 nM in agreement with previous reports (Linsley, P.S. et al. (1988) J. Virol 62:3695-3702; Schnittman, et al. (1988) J. Immunol. 141:4181-4186). Calculations from the association and dissociation rate constants gave a similar comparative result. The expressed gp120r has a relative molecular mass (Mr) of ~48,500 and a protein of similar size was also partially purified from placental membranes (FIGURE 2D).

The placental membranes and COS cells were surface iodinated, and treated with 1 nM unlabelled vgp120, then washed with Blotto RPMI, 5% BSA, 1% Non-fat dry milk, 0.2% sodium azide solubilized in Triton X-100 (1% in PBS with a protein inhibitor cocktail, PMSF, Pepstatin A, orthophenathroline and leupeptin) and immunoprecipitated

with HIV or control human sera, according to the method described in Curtis et al. (1990) J. Immunol. 144:1295-1303.

Northern analysis of the expression of the gp120r RNA indicated a major species of ~5 kb and a minor species of ~1.7 kb which may represent an alternatively processed transcript and is more consistent with the size of the gp120r cDNA. RNA was denatured, separated in an agarose gel, transferred to nitrocellulose, hybridized to gp120r cDNA and autoradiographed for 3 days.

Expression of gp120r RNA was highest in colon followed by thymus, placenta, heart, skeletal muscle, and was not detected in liver or kidney. Low levels of expression in brain, T cell, B cell, and macrophage (FIGURE 2E) require verification by polymerase chain reaction (PCR). Full length CD4 RNA was highest in thymus, T cell, and macrophage followed by placenta and colon (not shown).

The gp120r cDNA encodes a protein of 404 amino acids with a calculated Mr of 45,775 (FIGURE 3A).

Sequencing of both strands of gp120r cDNA was carried out by the dideoxy chain termination method. The nucleotide sequence preceeding the first ATG agrees with the Kozak consensus. The predicted cytoplasmic domain has a similar length and shows some sequence homology to other type II membrane protein C-type lectins (Spiess, M. (1990) Biochemistry 29:10009-10018). The membrane spanning sequence is underlined and was predicted in part by homology to related sequences in FIGURE 3C. The potential N-linked glycosylation site is marked by an asterisk. The start of the seven complete and eighth partial tandem repeats are indicated (R1-R8). The consensus repeat sequence is IYQELT(R/Q) LKAAVGELPEKSKLQE. The beginning of the lectin domains is also indicated (L). No signal sequence was apparent but instead demonstrated homology to a family of Type II membrane proteins which utilize a ~20 residue hydrophobic stop-transfer sequence for membrane translocation. The "positive inside rule" (von Heijne, G. et al. (1988) Eur. J. Biochem. 174:671-678) for the sequence within fifteen residues of the transmembrane region predicts a cytoplasmic amino terminus in agreement with the homology to membrane associated C-type lectins with similar membrane orientation (FIGURE 3C) (Spiess, M. (1990) Biochemistry 29:10009-10018). This region, Met 1 to Ala 76, represents the first domain of the gp120r sequence.

The second domain (Ile 77 to Val 249) consists of tandem repeats of nearly identical sequence (FIGURE 3A). This region was predicted to consist of a series of amphipathic α-helices interrupted by ß-turns. Circular Dichroism spectra in 40% trifluoroethanol of a consensus repeat peptide beginning with the ß-turn, PEKSKLQEIYQELTQLKAAVGEL (single-letter amino-acid code), demonstrated an all α-helical structure (not shown). Homology to other repeat domains suggested three possible

tertiary structures, (1) antiparallel helix bundles, (2) a multimeric parallel helix bundle, and (3) a membrane pore with a hydrophobic exterior and a negatively charged interior. The first two models would function as spacers to separate the lectin domain from the membrane, while the third could generate a transmembrane signal after ligand binding.

The third domain (Cys 253 to Ala 404) is homologous to the other known C-type lectins which are type II membrane proteins (FIGURE 3C). With the exception of the IgEr, these lectins bind terminal D-galactose and D-N-acetylgalactosamine of glycoproteins (Spiess, M. (1990) Biochemistry 29: 10009-10018).

The most closely related sequences were the group of Type II membrane protein C-type lectins: Chick hepatic lectin (CHL) (Drickamer, K.J. (1981) Biol. Chem. 256:5827-5839), low affinity IgE receptor (IgEr) (Kikutani, H. et al. (1986) Cell 47: 657-665), the asialoglycoptorein receptors (human H1 and H2 (Spiess, M. et al. (1985) Proc. Natl. Acad. Sci. USA 82:6465-6569) are shown), and the rat Kupffer cell receptor (Hoyle, G.W. et al. (1988) J. Biol. Chem. 263:7487-7492). The most similar mannose binding lectin was one of the eight carbohydrate recognition domains of the human macrophage mannose receptor (Mannr) (Taylor, M.E. et al. (1990) J. Biol. Chem. 265:12156-12162; Ezekowitz, R.A.B. et al. (1990) J. Exp. Med. 172:1785-1794). Residues identical to the gp120r are boxed. ALIGN scores indicate significant sequence similarity if greater than 3.0. The complete gp120r sequence was most homologous to the Kupffer cell receptor which has a similar tandem repeat (Hoyle, G.W. et al. (1988) J. Biol. Chem. 263:7487-7492).

The inability to crosslink gp120 to the non-CD4 sites on placenta and brain cell lines (not shown) was consistent with an interaction of the gp120r with carbohydrate, and polyclonal HIV antisera added to gp120 blocked binding to CD4 but not to the gp120r (FIGURE 1B). Galactose and N-acetylgalactosamine did not block gp120 binding, but mannose and fucose completely blocked binding to the gp120r without an effect on CD4 (FIGURE 1B). Inhibition by a series of sugars is shown in FIGURE 2B. Human IgE (10 μg/ml), sialic acid (100 mM), and mannose-6-phosphate (100 mM) had no effect on binding to the gp120r. The three forms of gp120 used have different oligosaccaride structures. Bgp120 contains only high mannose structures (Hsieh, P. et al. (1984) J. Biol. Chem. 259:2375-2382). Vgp120 has equal proportions of high mannose and complex (Mizuochi, T. et al. (1988) Biochem. J. 254:599-603) similar to ngp120 which has a greater structural diversity in the complex chains (Geyer, H. et al. (1988) J. Biol. Chem. 263:11760-11767; Mizuochi, T. et al. (1990) J. Biol. Chem. 265:8519-8524). The affinity of the gp120r for all three forms was similar (FIGURE 2A) suggesting that the terminal mannose of high mannose chains are the primary determinants of binding. As expected for a C-type lectin the gp120r required calcium and binding was blocked by

EGTA (FIGURE 2B). The gp120r carbohydrate specificity is more closely related to serum mannose binding proteins and to the Mr 175,000 mannose-specific endocytosis receptor found in macrophages and placenta (Taylor, M.E. et al. (1990) J. Biol. Chem. 265:12156-12162; Ezekowitz, R.A.B. et al. (1990) J. Exp. Med. 172:1785-1794) (FIGURE 3C). Low (1 nM) concentrations of gp120 did not purify a Mr 175,000 band from placental membranes (FIGURE 2D) consistent with a reported concentration of 150-300 nM for gp120 saturation of the macrophage receptor (Larkin, M. et al. (1989) AIDS 3, 793-798).

The importance of gp120 carbohydrate in HIV infection has been suggested by the ability of plant lectins (Lifson, J. et al. (1986) E. J. Exp. Med. 164:2101-2106) and serum mannose-binding protein (Ezekowitz, R.A.B. et al. (1989) J. Exp. Med. 169:185-196) to block infection, and a proposed role for the macrophage endocytosis receptor in viral attachment (Larking M. et al. (1989) AIDS 3, 793-798). Concanavalin A treatment of gp120 blocked binding to the gp120r and CD4 (FIGURE 1B), consistent with a steric hindrance of receptor interaction. The antibiotic pradimicin A blocks HIV infection of CD4 positive T cells and this inhibitory effect is prevented by mannan and EGTA (Tanabe-Tochikura. A. et al. (1990) Virology 176:476-473). Pradamicin blocked gp120 binding to the gp120r and CD4, while mannan and EGTA only inhibited binding to the gp120r (FIGURE2B). Mannan inhibited ~10% of high affinity (nM) gp120 binding to T cells and macrophages, consistent with gp120r expression (FIGURE 2E), suggesting that in addition to CD4 the gp120r may be important for HIV binding and infection. The observation the the gp120r rapidly internalized its bound ligand gp120 (FIGURE 2C), and also binds radiolabelled HIV in a gp120 dependent fashion (FIGURE 1C) also support this conclusion.

The foregoing description and Examples are intended as illustrative of the present invention, but not as limiting. Numerous variations and modifications may be effected without departing from the true spirit and scope of the present invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Curtis, Benson

(ii) TITLE OF INVENTION: INHIBITION OF NON-CD4 MEDIATED HIV INFECTION

(iii) NUMBER OF SEQUENCES: 9

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Bristol-Myers Squibb Company
(B) STREET: 3005 First Avenue
(C) CITY: Seattle
(D) STATE: Washington
(E) COUNTRY: USA
(F) ZIP: 98121

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US UNKNOWN
(B) FILING DATE: 11-JUL-1991
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Sorrentino, Joseph M.
(B) REGISTRATION NUMBER: 32,598
(C) REFERENCE/DOCKET NUMBER: ON0086-

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (206) 728-4800
(B) TELEFAX: (206) 448-4775

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1312 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Human immunodeficiency virus type 1

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 42..1253

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
CTAAAGCAGG AGTTCTGGAC ACTGGGGGAG AGTGGGGTGA C ATG AGT GAC TCC      53
                                              Met Ser Asp Ser
                                                1

AAG GAA CCA AGA CTG CAG CAG CTG GGC CTC CTG GAG GAG GAA CAG CTG    101
Lys Glu Pro Arg Leu Gln Gln Leu Gly Leu Leu Glu Glu Glu Gln Leu
  5                  10                  15                  20

AGA GGC CTT GGA TTC CGA CAG ACT CGA GGA TAC AAG AGC TTA GCA GGG    149
Arg Gly Leu Gly Phe Arg Gln Thr Arg Gly Tyr Lys Ser Leu Ala Gly
                 25                  30                  35

TGT CTT GGC CAT GGT CCC CTG GTG CTG CAA CTC CTC TCC TTC ACG CTC    197
Cys Leu Gly His Gly Pro Leu Val Leu Gln Leu Leu Ser Phe Thr Leu
             40                  45                  50

TTG GCT GGG CTC CTT GTC CAA GTG TCC AAG GTC CCC AGC TCC ATA AGT    245
Leu Ala Gly Leu Leu Val Gln Val Ser Lys Val Pro Ser Ser Ile Ser
         55                  60                  65

CAG GAA CAA TCC AGG CAA GAC GCG ATC TAC CAG AAC CTG ACC CAG CTT    293
Gln Glu Gln Ser Arg Gln Asp Ala Ile Tyr Gln Asn Leu Thr Gln Leu
     70                  75                  80

AAA GCT GCA GTG GGT GAG CTC TCA GAG AAA TCC AAG CTG CAG GAG ATC    341
Lys Ala Ala Val Gly Glu Leu Ser Glu Lys Ser Lys Leu Gln Glu Ile
 85                  90                  95                 100
```

```
TAC CAG GAG CTG ACC CAG CTG AAG GCT GCA GTG GGT GAG CTT CCA GAG     389
Tyr Gln Glu Leu Thr Gln Leu Lys Ala Ala Val Gly Glu Leu Pro Glu
                105                 110                 115

AAA TCT AAG CTG CAG GAG ATC TAC CAG GAG CTG ACC CGG CTG AAG GCT     437
Lys Ser Lys Leu Gln Glu Ile Tyr Gln Glu Leu Thr Arg Leu Lys Ala
            120                 125                 130

GCA GTG GGT GAG CTT CCA GAG AAA TCT AAG CTG CAG GAG ATC TAC CAG     485
Ala Val Gly Glu Leu Pro Glu Lys Ser Lys Leu Gln Glu Ile Tyr Gln
        135                 140                 145

GAG CTG ACC TGG CTG AAG GCT GCA GTG GGT GAG CTT CCA GAG AAA TCT     533
Glu Leu Thr Trp Leu Lys Ala Ala Val Gly Glu Leu Pro Glu Lys Ser
    150                 155                 160

AAG ATG CAG GAG ATC TAC CAG GAG CTG ACT CGG CTG AAG GCT GCA GTG     581
Lys Met Gln Glu Ile Tyr Gln Glu Leu Thr Arg Leu Lys Ala Ala Val
165                 170                 175                 180

GGT GAG CTT CCA GAG AAA TCT AAG CAG CAG GAG ATC TAC CAG GAG CTG     629
Gly Glu Leu Pro Glu Lys Ser Lys Gln Gln Glu Ile Tyr Gln Glu Leu
                185                 190                 195

ACC CGG CTG AAG GCT GCA GTG GGT GAG CTT CCA GAG AAA TCT AAG CAG     677
Thr Arg Leu Lys Ala Ala Val Gly Glu Leu Pro Glu Lys Ser Lys Gln
            200                 205                 210

CAG GAG ATC TAC CAG GAG CTG ACC CGG CTG AAG GCT GCA GTG GGT GAG     725
Gln Glu Ile Tyr Gln Glu Leu Thr Arg Leu Lys Ala Ala Val Gly Glu
        215                 220                 225

CTT CCA GAG AAA TCT AAG CAG CAG GAG ATC TAC CAG GAG CTG ACC CAG     773
Leu Pro Glu Lys Ser Lys Gln Gln Glu Ile Tyr Gln Glu Leu Thr Gln
    230                 235                 240

CTG AAG GCT GCA GTG GAA CGC CTG TGC CAC CCC TGT CCC TGG GAA TGG     821
Leu Lys Ala Ala Val Glu Arg Leu Cys His Pro Cys Pro Trp Glu Trp
245                 250                 255                 260

ACA TTC TTC CAA GGA AAC TGT TAC TTC ATG TCT AAC TCC CAG CGG AAC     869
Thr Phe Phe Gln Gly Asn Cys Tyr Phe Met Ser Asn Ser Gln Arg Asn
                265                 270                 275

TGG CAC GAC TCC ATC ACC GCC TGC AAA GAA GTG GGG GCC CAG CTC GTC     917
Trp His Asp Ser Ile Thr Ala Cys Lys Glu Val Gly Ala Gln Leu Val
            280                 285                 290

GTA ATC AAA AGT GCT GAG GAG CAG AAC TTC CTA CAG CTG CAG TCT TCC     965
Val Ile Lys Ser Ala Glu Glu Gln Asn Phe Leu Gln Leu Gln Ser Ser
        295                 300                 305
```

AGA AGT AAC CGC TTC ACC TGG ATG GGA CTT TCA GAT CTA AAT CAG GAA 1013
Arg Ser Asn Arg Phe Thr Trp Met Gly Leu Ser Asp Leu Asn Gln Glu
310 315 320

GGC ACG TGG CAA TGG GTG GAC GGC TCA CCT CTG TTG CCC AGC TTC AAG 1061
Gly Thr Trp Gln Trp Val Asp Gly Ser Pro Leu Leu Pro Ser Phe Lys
325 330 335 340

CAG TAT TGG AAC AGA GGA GAG CCC AAC AAC GTT GGG GAG GAA GAC TGC 1109
Gln Tyr Trp Asn Arg Gly Glu Pro Asn Asn Val Gly Glu Glu Asp Cys
345 350 355

GCG GAA TTT AGT GGC AAT GGC TGG AAC GAC GAC AAA TGT AAT CTT GCC 1157
Ala Glu Phe Ser Gly Asn Gly Trp Asn Asp Asp Lys Cys Asn Leu Ala
360 365 370

AAA TTC TGG ATC TGC AAA AAG TCC GCA GCC TCC TGC TCC AGG GAT GAA 1205
Lys Phe Trp Ile Cys Lys Lys Ser Ala Ala Ser Cys Ser Arg Asp Glu
375 380 385

GAA CAG TTT CTT TCT CCA GCC CCT GCC ACC CCA AAC CCC CCT CCT GCG 1253
Glu Gln Phe Leu Ser Pro Ala Pro Ala Thr Pro Asn Pro Pro Pro Ala
390 395 400

TAGCAGAACT TCACCCCCTT TTAAGCTACA GTTCCTTCTC TCCATCCTTC GACCTTTAG 1312

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 404 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ser Asp Ser Lys Glu Pro Arg Leu Gln Gln Leu Gly Leu Leu Glu
1 5 10 15

Glu Glu Gln Leu Arg Gly Leu Gly Phe Arg Gln Thr Arg Gly Tyr Lys
20 25 30

Ser Leu Ala Gly Cys Leu Gly His Gly Pro Leu Val Leu Gln Leu Leu
35 40 45

Ser Phe Thr Leu Leu Ala Gly Leu Leu Val Gln Val Ser Lys Val Pro
50 55 60

Ser Ser Ile Ser Gln Glu Gln Ser Arg Gln Asp Ala Ile Tyr Gln Asn
65 70 75 80

Leu Thr Gln Leu Lys Ala Ala Val Gly Glu Leu Ser Glu Lys Ser Lys
85 90 95

Leu Gln Glu Ile Tyr Gln Glu Leu Thr Gln Leu Lys Ala Ala Val Gly
100 105 110

Glu Leu Pro Glu Lys Ser Lys Leu Gln Glu Ile Tyr Gln Glu Leu Thr
115 120 125

Arg Leu Lys Ala Ala Val Gly Glu Leu Pro Glu Lys Ser Lys Leu Gln
130 135 140

Glu Ile Tyr Gln Glu Leu Thr Trp Leu Lys Ala Ala Val Gly Glu Leu
145 150 155 160

Pro Glu Lys Ser Lys Met Gln Glu Ile Tyr Gln Glu Leu Thr Arg Leu
165 170 175

Lys Ala Ala Val Gly Glu Leu Pro Glu Lys Ser Lys Gln Gln Glu Ile
180 185 190

Tyr Gln Glu Leu Thr Arg Leu Lys Ala Ala Val Gly Glu Leu Pro Glu
195 200 205

Lys Ser Lys Gln Gln Glu Ile Tyr Gln Glu Leu Thr Arg Leu Lys Ala
210 215 220

Ala Val Gly Glu Leu Pro Glu Lys Ser Lys Gln Gln Glu Ile Tyr Gln
225 230 235 240

Glu Leu Thr Gln Leu Lys Ala Ala Val Glu Arg Leu Cys His Pro Cys
245 250 255

Pro Trp Glu Trp Thr Phe Phe Gln Gly Asn Cys Tyr Phe Met Ser Asn
260 265 270

Ser Gln Arg Asn Trp His Asp Ser Ile Thr Ala Cys Lys Glu Val Gly
275 280 285

Ala Gln Leu Val Val Ile Lys Ser Ala Glu Glu Gln Asn Phe Leu Gln
290 295 300

Leu Gln Ser Ser Arg Ser Asn Arg Phe Thr Trp Met Gly Leu Ser Asp
305 310 315 320

Leu Asn Gln Glu Gly Thr Trp Gln Trp Val Asp Gly Ser Pro Leu Leu
325 330 335

Pro Ser Phe Lys Gln Tyr Trp Asn Arg Gly Glu Pro Asn Asn Val Gly
340 345 350

Glu Glu Asp Cys Ala Glu Phe Ser Gly Asn Gly Trp Asn Asp Asp Lys
355 360 365

Cys Asn Leu Ala Lys Phe Trp Ile Cys Lys Lys Ser Ala Ala Ser Cys
370 375 380

Ser Arg Asp Glu Glu Gln Phe Leu Ser Pro Ala Pro Ala Thr Pro Asn
385 390 395 400

Pro Pro Pro Ala

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 127 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Human immunodeficiency virus type 1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Cys His Pro Cys Pro Trp Glu Trp Thr Phe Phe Gln Gly Asn Cys Tyr
1 5 10 15

Phe Met Ser Asn Ser Gln Arg Asn Trp His Asp Ser Ile Thr Ala Cys
20 25 30

Lys Glu Val Gly Ala Gln Leu Val Val Ile Lys Ser Ala Glu Glu Gln
35 40 45

Asn Phe Leu Gln Leu Gln Ser Ser Arg Ser Asn Arg Phe Thr Trp Met
50 55 60

Gly Leu Ser Asp Leu Asn Gln Glu Gly Thr Trp Gln Trp Val Asp Gly
65 70 75 80

Ser Pro Leu Leu Pro Ser Phe Lys Gln Tyr Trp Asn Arg Gly Glu Pro
85 90 95

Asn Asn Val Gly Glu Glu Asp Cys Ala Glu Phe Ser Gly Asn Gly Trp
100 105 110

Asn Asp Asp Lys Cys Asn Leu Ala Lys Phe Trp Ile Cys Lys Lys
115 120 125

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	126 amino acids
(B)	TYPE:	amino acid
(D)	TOPOLOGY:	linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Cys Gly Ala Gln Ser Arg Gln Trp Glu Tyr Phe Glu Gly Arg Cys Tyr
1 5 10 15

Tyr Phe Ser Leu Ser Arg Met Ser Trp His Lys Ala Lys Ala Glu Cys
20 25 30

Glu Glu Met His Ser His Leu Ile Ile Ile Asp Ser Tyr Ala Lys Gln
35 40 45

Asn Phe Val Met Phe Arg Thr Arg Asn Glu Arg Phe Trp Ile Gly Leu
50 55 60

Thr Asp Glu Asn Gln Glu Gly Glu Trp Gln Trp Val Asp Gly Thr Asp
65 70 75 80

Thr Arg Ser Ser Phe Thr Phe Trp Lys Glu Gly Glu Pro Asn Asn Arg
85 90 95

Gly Phe Asn Glu Asp Cys Ala His Val Trp Thr Ser Gly Gln Trp Asn
100 105 110

Asp Val Tyr Cys Thr Tyr Glu Cys Tyr Tyr Val Cys Glu Lys
115 120 125

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 125 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
Cys Asn Thr Cys Pro Glu Lys Trp Ile Asn Phe Gln Arg Lys Cys Tyr
1               5                   10                  15

Tyr Phe Gly Lys Gly Thr Lys Gln Trp Val His Ala Arg Tyr Ala Cys
            20                  25                  30

Asp Asp Met Glu Gly Gln Leu Val Ser Ile His Ser Pro Glu Glu Gln
        35                  40                  45

Asp Phe Leu Thr Lys His Ala Ser His Thr Gly Ser Trp Ile Gly Leu
    50                  55                  60

Arg Asn Leu Asp Leu Lys Gly Glu Phe Ile Trp Val Asp Gly Ser His
65                  70                  75                  80

Val Asp Tyr Ser Asn Trp Ala Pro Gly Glu Pro Thr Ser Arg Ser Gln
                85                  90                  95

Gly Glu Asp Cys Val Met Met Arg Gly Ser Gly Arg Trp Asn Asp Ala
            100                 105                 110

Phe Cys Asp Arg Lys Leu Gly Ala Trp Val Cys Asp Arg
        115                 120                 125
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 129 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
Arg Thr Cys Cys Pro Val Asn Trp Val Glu His Glu Arg Ser Cys Tyr
1               5                   10                  15

Trp Phe Ser Arg Ser Gly Lys Ala Trp Ala Asp Ala Asp Asn Tyr Cys
            20                  25                  30

Arg Leu Glu Asp Ala His Leu Val Val Val Thr Ser Trp Glu Glu Gln
        35                  40                  45

Lys Phe Val Gln His His Ile Gly Pro Val Asn Thr Trp Met Gly Leu
    50                  55                  60

His Asp Gln Asn Gly Pro Trp Lys Trp Val Asp Gly Thr Asp Tyr Glu
65                  70                  75                  80

Thr Gly Phe Lys Asn Trp Arg Pro Glu Gln Pro Asp Asp Trp Tyr Gly
                85                  90                  95

His Gly Leu Gly Gly Gly Glu Asp Cys Ala His Phe Thr Asp Asp Gly
            100                 105                 110

Arg Trp Asn Asp Asp Val Cys Gln Arg Pro Tyr Arg Trp Val Cys Glu
        115                 120                 125

Thr
```

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 129 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```
Arg Thr Cys Cys Pro Val Asn Trp Val Glu His Gln Gly Ser Cys Tyr
1               5                   10                  15

Trp Phe Ser His Ser Gly Lys Ala Trp Ala Glu Ala Glu Lys Tyr Cys
            20                  25                  30

Gln Leu Glu Asn Ala His Leu Val Val Ile Asn Ser Trp Glu Glu Gln
        35                  40                  45

Lys Phe Ile Val Gln His Thr Asn Pro Phe Asn Thr Trp Ile Gly Leu
    50                  55                  60

Thr Asp Ser Asp Gly Ser Trp Lys Trp Val Asp Gly Thr Asp Tyr Arg
65                  70                  75                  80

His Asn Tyr Lys Asn Trp Ala Val Thr Gln Pro Asp Asn Trp His Gly
                85                  90                  95

His Glu Leu Gly Gly Ser Glu Asp Cys Val Glu Val Gln Pro Asp Gly
            100                 105                 110

Arg Trp Asn Asp Asp Phe Cys Leu Gln Val Tyr Arg Trp Val Cys Glu
        115                 120                 125

Lys
```

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 130 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```
Leu Gln Leu Ile Met Gln Asp Trp Lys Tyr Phe Asn Gly Lys Phe Tyr
1               5                   10                  15
```

```
Tyr Phe Ser Arg Asp Lys Lys Ser Trp His Glu Ala Glu Asn Phe Cys
            20                  25                  30

Val Ser Gln Gly Ala His Leu Ala Ser Val Thr Ser Gln Glu Glu Gln
        35                  40                  45

Ala Phe Leu Val Gln Ile Thr Asn Ala Val Asp His Trp Ile Gly Leu
    50                  55                  60

Thr Asp Gln Gly Thr Glu Gly Asn Trp Arg Trp Val Asp Gly Thr Pro
65                  70                  75                  80

Phe Asp Tyr Val Gln Ser Arg Arg Phe Trp Arg Lys Gly Gln Pro Asp
                85                  90                  95

Asn Trp Arg His Gly Asn Gly Glu Arg Glu Asp Cys Val His Leu Gln
            100                 105                 110

Arg Met Trp Asn Asp Met Ala Cys Gly Thr Ala Tyr Asn Trp Val Cys
        115                 120                 125

Lys Lys
    130
```

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 130 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

```
Pro Thr His Cys Pro Ser Gln Trp Trp Pro Tyr Ala Gly His Cys Tyr
1               5                   10                  15

Lys Ile His Arg Asp Glu Lys Lys Ile Gln Arg Asp Ala Leu Thr Thr
            20                  25                  30

Cys Arg Lys Glu Gly Gly Asp Leu Thr Ser Ile His Thr Ile Glu Glu
        35                  40                  45
```

Leu Asp Phe Ile Ile Ser Gln Leu Gly Leu Glu Pro Asn Asp Glu Leu
50 55 60

Trp Ile Gly Leu Asn Asp Ile Lys Ile Gln Met Tyr Phe Glu Trp Ser
65 70 75 80

Asp Gly Thr Pro Val Thr Phe Thr Lys Trp Leu Arg Gly Glu Pro Ser
85 90 95

His Glu Asn Asn Arg Gln Glu Asp Cys Val Val Met Lys Gly Lys Asp
100 105 110

Gly Tyr Trp Ala Asp Arg Gly Cys Glu Trp Pro Leu Gly Tyr Ile Cys
115 120 125

Lys Met
130

We claim:

1. A method of inhibiting HIV infection of mammalian cells comprising contacting the cells with an effective amount of a compound selected from the group consisting of a mannose carbohydrate, a fucose carbohydrate, a lectin and a drug, for a time period sufficient to significantly inhibit the binding of HIV to a non-CD4 cell surface protein.

2. The method of Claim 1, wherein the non-CD4 cell surface protein is a gp120 receptor having a specific binding affinity for gp120 of about Kd = 1.3 nM to about Kd = 2.0 nM.

3. The method of Claim 2, wherein the gp120 receptor is present on placental cells.

4. The method of Claim 2, wherein the gp120 receptor is present on muscle cells.

5. The method of Claim 2, wherein the gp120 receptor is present on neural cells.

6. The method of Claim 5, wherein the neural cells are brain cells.

7. The method of Claim 5, wherein the neural cells are dendritic cells.

8. The method of Claim 2, wherein the gp120 receptor is present on mucosal cells.

10. The method of Claim 1, wherein the compound is mannose.

10. The method of Claim 1, wherein the compound is fucose.

11. The method of Claim 1, wherein the compound is a mannose-containing carbohydrate.

12. The method of Claim 11, where the carbohydrate is mannan.

13. The method of Claim 1, wherein the compound is a pradimicin A antibiotic.

14. A substantially purified non-CD4 gp120 receptor protein comprising a protein substantially corresponding to a non-CD4 mammalian cell surface protein that has a

specific binding affinity for gp120, said protein containing about 400 amino acid residues, having a molecular weight of about 45,000 daltons and having a binding affinity for gp120 characterized by a Kd of about 1.3 nM to about 2 nM.

15. The gp120 receptor protein of Claim 14, wherein the binding of the gp120 receptor protein to gp120 is inhibited by a compound selected from the group consisting of a mannose carbohydrate, a fucose carbohydrate, a lectin and a drug.

16. The gp120 receptor of Claim 15, wherein the compound is mannose.

17. The gp120 receptor protein of Claim 15, wherein the compound is a pradimicin A antibiotic.

18. The gp120 receptor protein of Claim 14, wherein the protein is produced by recombinant means.

19. The gp120 receptor protein of Claim 18, wherein said recombinant means comprises the cloning of a cDNA isolated from a library of recombinant placental genes.

20. A DNA molecule encoding the gp120 receptor protein of Claim 14, wherein the DNA is a complementary DNA that transcribes an mRNA found in cells selected from the group consisting of placental cells, brain cells, muscle cells and colon cells.

21. A method of detecting the presence of HIV in a sample comprising:

(a) admixing in an aqueous medium a sample to be assayed with a non-CD4 gp120 receptor protein having a specific binding affinity for gp120 characterized by a Kd of about 1.3 nM to about 2.0 nM in an amount sufficient to carry out at least one assay;

(b) maintaining the admixture for a time period sufficient for the gp120 receptor protein to bind to any HIV present in the sample and form a reaction product; and

(c) determining the presence of the HIV containing reaction product.

22. The method of Claim 21, wherein the gp120 receptor protein contains about 400 amino acid residues and has a molecular weight of about 45,000 daltons.

23. The method of Claim 21, wherein the gp120 receptor protein is affixed to a solid matrix to form a solid support.

24. The method of Claim 21, wherein the presence of the reaction product is determined by contacting the sample with a reagent capable of detecting the bound gp120 receptor protein.

25. The method of Claim 24, wherein the reagent is a labelled antibody directed against the gp120 receptor protein.

26. A diagnostic system in kit form, for assaying for the presence of HIV in a fluid sample, comprising a package containing a non-CD4 receptor protein having a specific affinity for gp120 characterized by a Kd of about 1.3 nM to about 2.0 nM, and instructions for use.

27. The diagnostic system of Claim 26, wherein the non-CD4 gp120 receptor protein is affixed to a solid matrix to form a solid support.

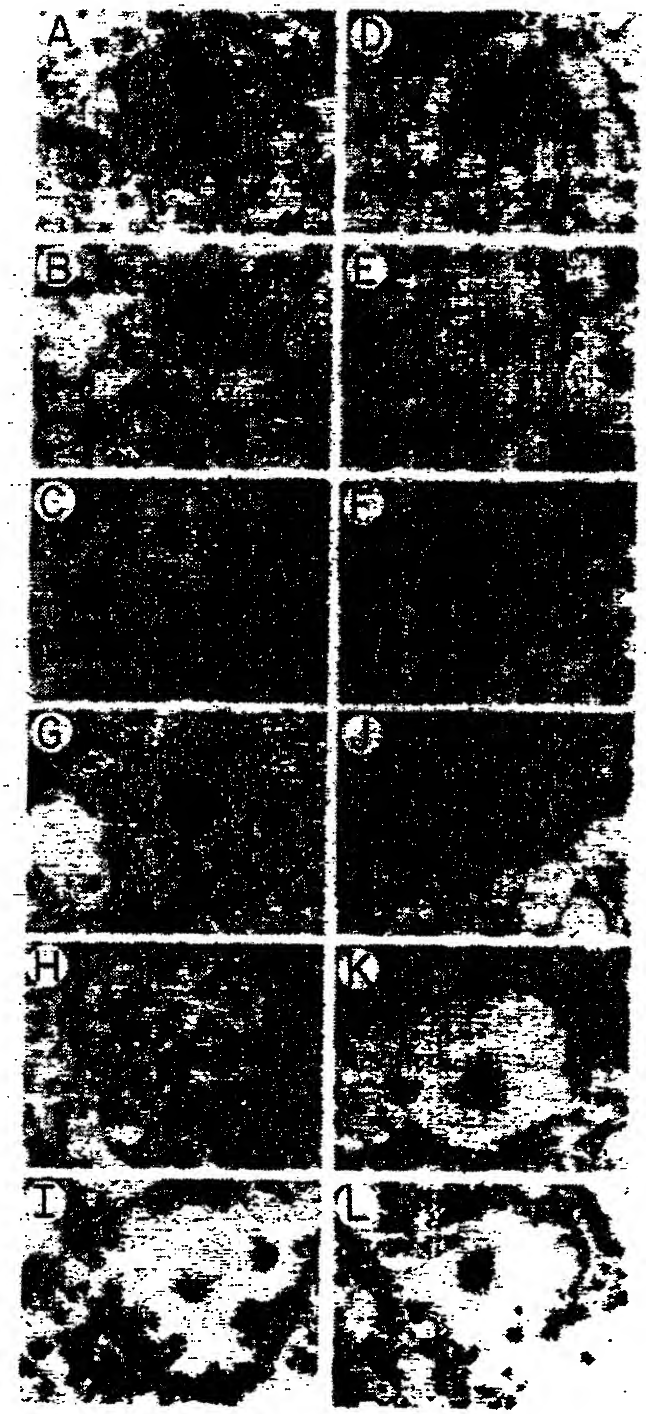


Figure 1A

SUBSTITUTE SHEET

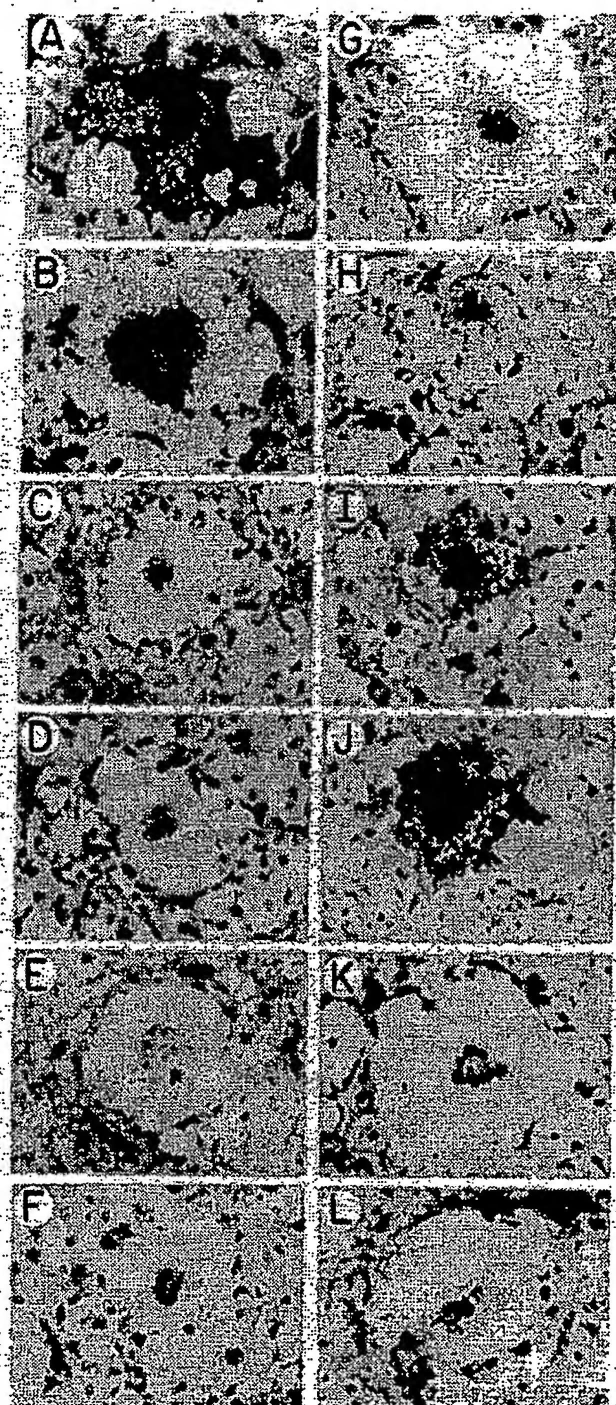


Figure 1B

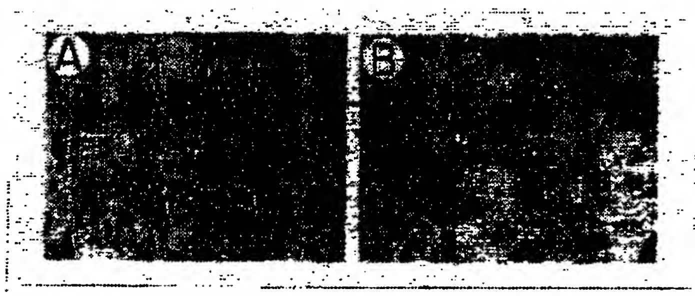


Figure 1C

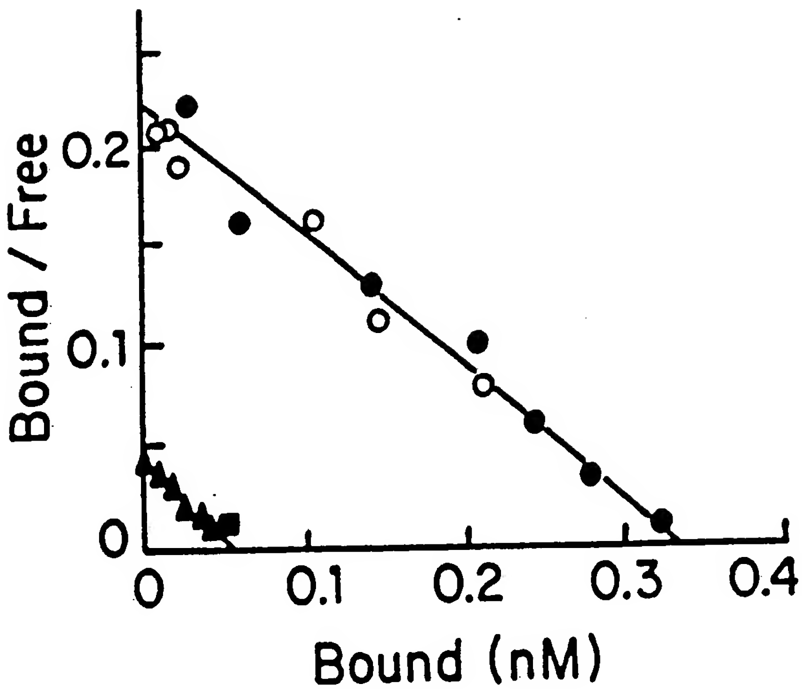
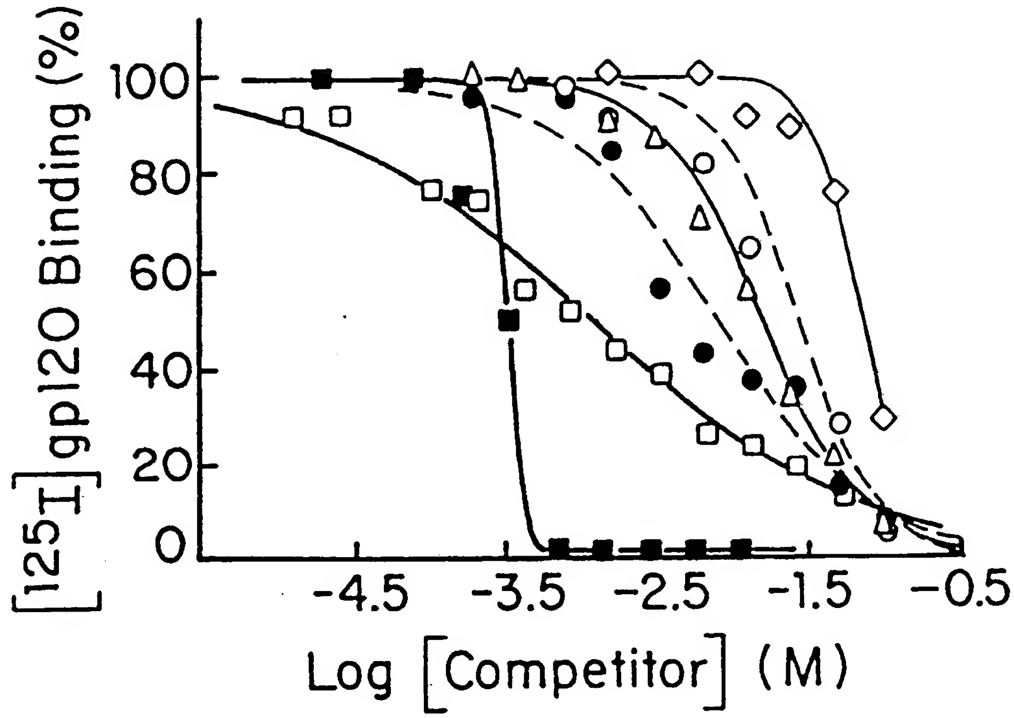


Figure 2A



[125I]gp120 Binding (%)
100
80
60
40
20
0
-4.5
-3.5
-2.5
-1.5
-0.5
Log [Competitor] (M)

Figure 2B

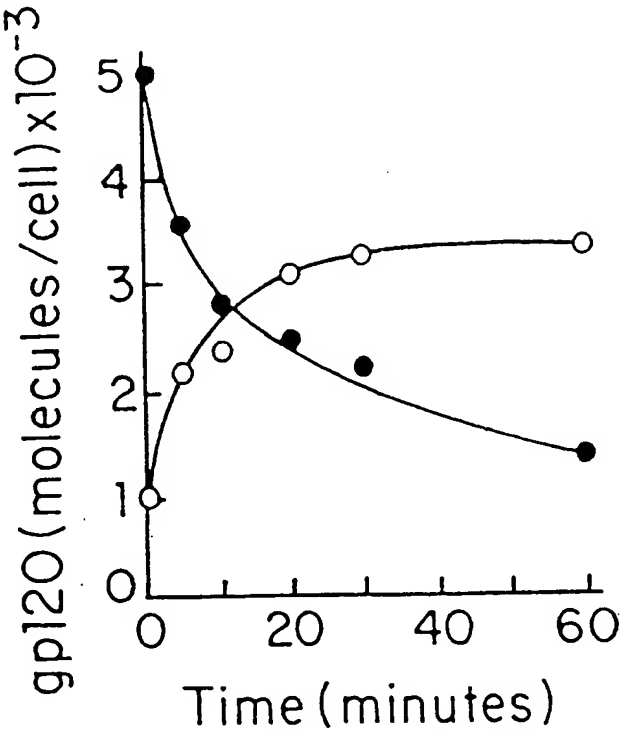


Figure 2C

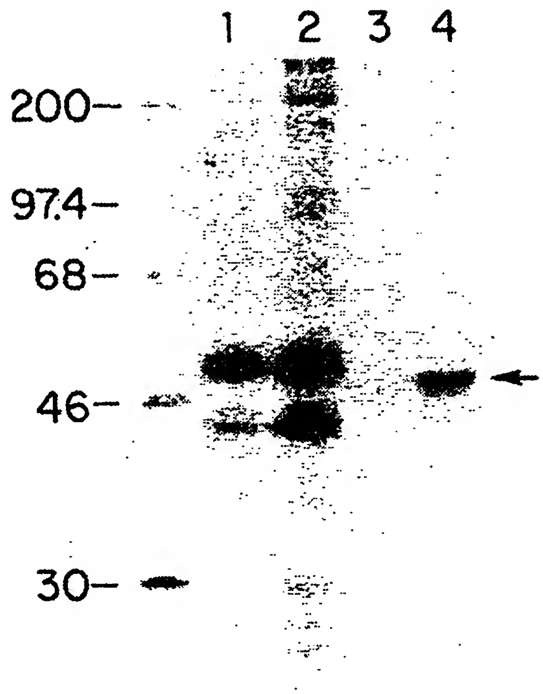


Figure 2D

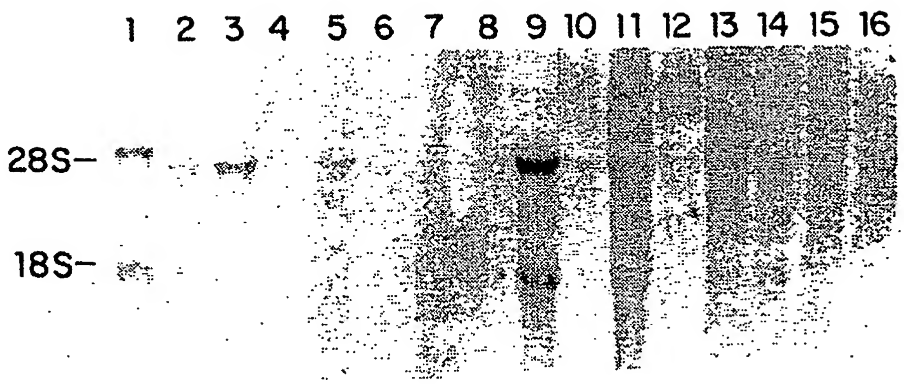


Figure 2E

```
  1                    CTAAAGCAGGAGTTCTGGACACTGGGGGAGAGTGGGGTGAC

 42   ATGAGTGACTCCAAGGAACCAAGACTGCAGCAGCTGGGCCTCCTGGAGGAGGAACAGCTG
  1   M  S  D  S  K  E  P  R  L  Q  Q  L  G  L  L  E  E  E  Q  L

102   AGAGGCCTTGGATTCCGACAGACTCGAGGATACAAGAGCTTAGCAGGGTGTCTTGGCCAT
 21   R  G  L  G  F  R  Q  T  R  G  Y  K  S  L  A  G  C  L  G  H

162   GGTCCCCTGGTGCTGCAACTCCTCTCCTTCACGCTCTTGGCTGGGCTCCTTGTCCAAGTG
 41   G  P  L  V  L  Q  L  L  S  F  T  L  L  A  G  L  L  V  Q  V

222   TCCAAGGTCCCCAGCTCCATAAGTCAGGAACAATCCAGGCAAGACGCGATCTACCAGAAC
 61   S  K  V  P  S  S  I  S  Q  E  Q  S  R  Q  D  A  I  Y  Q  N

                                                     R1       *
282   CTGACCCAGCTTAAAGCTGCAGTGGGTGAGCTCTCAGAGAAATCCAAGCTGCAGGAGATC
 81   L  T  Q  L  K  A  A  V  G  E  L  S  E  K  S  K  L  Q  E  I

                                                              R2
342   TACCAGGAGCTGACCCAGCTGAAGGCTGCAGTGGGTGAGCTTCCAGAGAAATCTAAGCTG
101   Y  Q  E  L  T  Q  L  K  A  A  V  G  E  L  P  E  K  S  K  L

402   CAGGAGATCTACCAGGAGCTGACCCGGCTGAAGGCTGCAGTGGGTGAGCTTCCAGAGAAA
121   Q  E  I  Y  Q  E  L  T  R  L  K  A  A  V  G  E  L  P  E  K

            R3
462   TCTAAGCTGCAGGAGATCTACCAGGAGCTGACCTGGCTGAAGGCTGCAGTGGGTGAGCTT
141   S  K  L  Q  E  I  Y  Q  E  L  T  W  L  K  A  A  V  G  E  L

                    R4
522   CCAGAGAAATCTAAGATGCAGGAGATCTACCAGGAGCTGACTCGGCTGAAGGCTGCAGTG
161   P  E  K  S  K  M  Q  E  I  Y  Q  E  L  T  R  L  K  A  A  V

                                 R5
582   GGTGAGCTTCCAGAGAAATCTAAGCAGCAGGAGATCTACCAGGAGCTGACCCGGCTGAAG
181   G  E  L  P  E  K  S  K  Q  Q  E  I  Y  Q  E  L  T  R  L  K

                                             R6
642   GCTGCAGTGGGTGAGCTTCCAGAGAAATCTAAGCAGCAGGAGATCTACCAGGAGCTGACC
201   A  A  V  G  E  L  P  E  K  S  K  Q  Q  E  I  Y  Q  E  L  T

                                                       R7
702   CGGCTGAAGGCTGCAGTGGGTGAGCTTCCAGAGAAATCTAAGCAGCAGGAGATCTACCAG
221   R  L  K  A  A  V  G  E  L  P  E  K  S  K  Q  Q  E  I  Y  Q

                                                          R8
762   GAGCTGACCCAGCTGAAGGCTGCAGTGGAACGCCTGTGCCACCCCTGTCCCTGGGAATGG
241   E  L  T  Q  L  K  A  A  V  E  R  L  C  H  P  C  P  W  E  W
                                         L

822   ACATTCTTCCAAGGAAACTGTTACTTCATGTCTAACTCCCAGCGGAACTGGCACGACTCC
261   T  F  F  Q  G  N  C  Y  F  M  S  N  S  Q  R  N  W  H  D  S
```

Figure 3A

```
 882  ATCACCGCCTGCAAAGAAGTGGGGGCCCAGCTCGTCGTAATCAAAAGTGCTGAGGAGCAG
 281  I  T  A  C  K  E  V  G  A  Q  L  V  V  I  K  S  A  E  E  Q

 942  AACTTCCTACAGCTGCAGTCTTCCAGAAGTAACCGCTTCACCTGGATGGGACTTTCAGAT
 301  N  F  L  Q  L  Q  S  S  R  S  N  R  F  T  W  M  G  L  S  D

1002  CTAAATCAGGAAGGCACGTGGCAATGGGTGGACGGCTCACCTCTGTTGCCCAGCTTCAAG
 321  L  N  Q  E  G  T  W  Q  W  V  D  G  S  P  L  L  P  S  F  K

1062  CAGTATTGGAACAGAGGAGAGCCCAACAACGTTGGGGAGGAAGACTGCGCGGAATTTAGT
 341  Q  Y  W  N  R  G  E  P  N  N  V  G  E  E  D  C  A  E  F  S

1122  GGCAATGGCTGGAACGACGACAAATGTAATCTTGCCAAATTCTGGATCTGCAAAAAGTCC
 361  G  N  G  W  N  D  D  K  C  N  L  A  K  F  W  I  C  K  K  S

1182  GCAGCCTCCTGCTCCAGGGATGAAGAACAGTTTCTTTCTCCAGCCCCTGCCACCCCAAAC
 381  A  A  S  C  S  R  D  E  E  Q  F  L  S  P  A  P  A  T  P  N

1242  CCCCCTCCTGCGTAGCAGAACTTCACCCCCTTTTAAGCTACAGTTCCTTCTCTCCATCCT
 401  P  P  P  A  ***

1302  TCGACCTTTAG
```

Figure 3A (cont.)

SUBSTITUTE SHEET

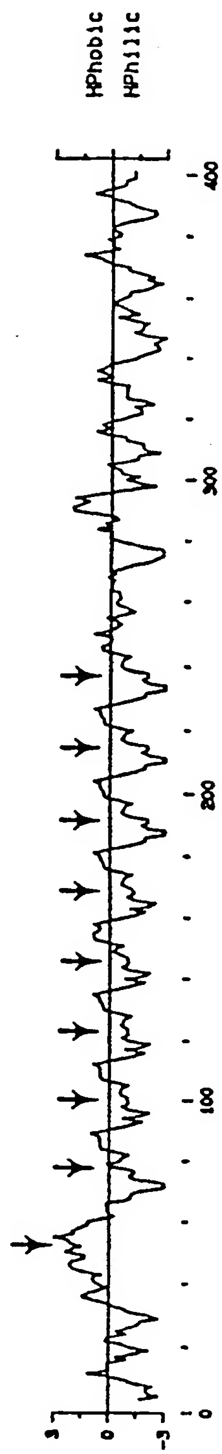


Figure 3B

```
gp120(253)CHPCEWEWTFFQGNCYFMSNS-QRNWHDSITACKEVGAQLVVIKSAEEQNFLQLQ-SSRSNRFTWMG
CHL   (78)CGAQSRQWEYFEGRCYYFSLS-RMSWHKAKAECEEMHSHLIIIDSYAKQNFVMFR--TRNERF-WIG
IgER (160)CNTCPEKWINFQRKCYYFGKG-TKQWVHARYACDDMEGQLVSIHSPEEQDFLTKH-ASHTG--SWIG
H1   (148)RTCCPVNWVEHERSCYWFSRS-GKAWADADNYCRLEDAHLVVVTSWEEQKFVQHH-IGPVN--TWMG
H2   (174)RTCCPVNWVEHQGSCYWFSHS-GKAWAEAEKYCQLENAHLVVINSWEEQKFIVQH-TNPFN--TWIG
Kupff(409)LQLIMQDWKYFNGKFYYFSRD-KKSWHEAENFCVSQGAHLASVTSQEEQAFLVQI-TNAVDH--WIG
Mannr(341)PTHCPSQWWPYAGHCYKIHRDEKKIQRDALTTCRKEGGDLTSIHTIEELDFIISQLGLEPNDELWIG

LSDLNQEGTWQWVDGSPL-LPSFKQYWNRGEPNNV------GEEDCAEFSGN-G-WNDDKCNLAKF-WICKK ALIGN
LTDENQEGEWQWVDGTDT-RSSFT-FWKEGEPNNR-----GFNEDCAHVWTS-GQWNDVYCTYECY-YVCEK 17.5
LRNLDLKGEFIWVDGSHV---DYSN-WAPGEPTSRS-----QGEDCVMMRGS-GRWNDAFCDRKLGAWVCDR 15.6
LHD--QNGPWKWVDGTDY-ETGFKN-WRPEQPDDWYGHGLGGGEDCAHFTDD-GRWNDDVC-QRPYRWVCET 13.1
LTD--SDGSWKWVDGTDY-RHNYKN-WAVTQPDNWHGHELGGSEDCVEVQPD-GRWNDDFC-LQVYRWVCEK 12.9
LTDQGTEGNWRWVDGTPFDYVQSRRFWRKGQPDNWRHGN-GEREDCVHLQRM---WNDMACGTA-YNWVCKK 13.2
LNDIKIQMYFEWSDGTPV---TFTK-WLRGEPSHENN----RQEDCVVMKGKDGYWADRGCEWPLG-YICKM 11.6
```

Figure 3C